U

Universal Multimedia Access

Definition

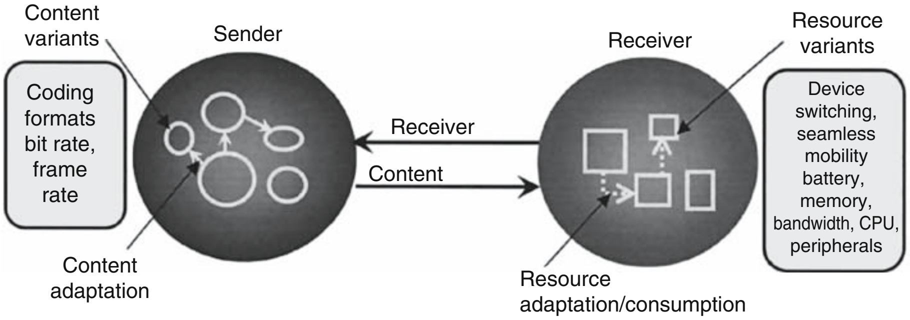
Universal multimedia access refers to access to multimedia content over wired and wireless networks on a range of devices with varying capabilities.

Recent technology advances have made possible access to digital multimedia content over wired and wireless networks on a range of devices with varying capabilities such as mobile phones, personal computers, and digital video recorders. This Universal Multimedia Access (UMA), enabled by new technologies and standards, poses new challenges and requires new solutions. Content delivery services to resource constrained devices such as mobile phones are limited due to the mismatch between the resources required to play the content and the device capabilities.

Figure 1 shows the key elements of a pervasive media delivery environment. The capabilities of the receivers in such an environment vary, requiring a server that can satisfy receivers with different capability sets. The capabilities of these devices would also change with the available battery, concurrently running applications, and available resources such as memory, bandwidth, and peripherals. As the available resources on a device change, the capability of the devices to process/playback content also changes. The resource variants shown in Fig. 1 represent the same device with changing capabilities. As the capabilities of a device change, the device cannot continue to play the same content. The content now has to be adapted to meet the resource requirements or the session has to be terminated. The content available at a sender may be adapted dynamically to meet the changing resource capabilities or a discrete number of content variants could be created offline to serve the receivers. The primary goal of content adaptation is to maximize the end users quality of experience given the resource constraints at the receiver and the sender.

Content Adaptation

The mismatch between the content and resources required to play the contest is bridged using adaptation techniques. The adaptation techniques can either adapt the content to match the receiver capabilities or adapt the resources to match the content. Resource adaptation typically takes the form of resource acquisition. The content adaptation problem has two aspects (1) determining what information to send and (2) how to encode that information efficiently for transmission. Determining the right information to be sent based on user preferences and available resources uses summarization techniques that strive to maintain semantic equivalence with minimal amount of information [1]. Once the appropriate



Universal Multimedia Access. Figure 1. Pervasive media delivery environment.

content has been determined, a suitable compression technique is selected and the content is transcoded to match the receiver capabilities [2].

Resource Adaptation

Another approach to UMA is through resource acquisition. A receiver acquires additional resources to bridge the mismatch with the content, primarily by collaborating with other devices in its environment and thus creating a virtual device. This virtual device approach to receiver adaptation was reported in [3]. The availability of Bluetooth and the upcoming short range ultra wideband communications make the virtual device a possibility. The virtual device concept works well in home and office environments with access to a number of peripherals. The key issue here is security: how can trust be established in a peripheral device.

Standardization

The UMA is also supported by the international standardization activities. The MPEG committee, under its MPEG-21 activity, has standardized tools for digital item adaptation (DIA) [4]. Digital item is a generic term for digital information that is exchanged between devices. The standard specifies tools for describing the digital items and adaptation alternatives when adaptation is necessary. The TV anytime forum has also released a series of specification to enable pervasive audio visual services (www.tv-anytime.org). The W3C has developed a standard to describe device capabilities and user preferences called Composite Capabilities/Preferences Profile (CC/PP) [5]. The CC/PP descriptions can be used during session setup to understand the receiver preferences and capabilities. These CC/PP descriptions can drive the content and resource adaptation necessary to make the content delivery possible.

References

1. L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-Summarization of Audio-Video Presentations," Proceedings of the Seventh ACM International Conference on Multimedia (Part 1), pp. 489–498, 1999.
2. J. Xin, C.-W. Lin, and M.-T. Sun, "Digital Video Transcoding," Proceedings of the IEEE, Vol. 93, No. 1, January 2005, pp. 84–97.
3. R.Y. Fu et al., "A Framework for Device Capability on Demand and Virtual Device User Experience," IBM Journal of Research and Development, Vol. 48, No. 5/6, 2004, pp. 635–648.
4. ISO/IEC JTC 1/SC 29, "Information technology – Multimedia framework (MPEG-21) – Part 7: Digital Item Adaptation?" ISO/IEC 21000-7:2004.
5. G. Klyne, et al., "Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies," W3C Recommendation, Available at: http://www.w3.org/TR/CCPP-struct-vocab/, 2004.

Usage Environment Adaptation

Synonyms

► Network and terminal adaptation

Definition

Usage environment adaptation refers to customization of network and terminal resources based on usage and content.

Usage environment adaptation refers to the case where instead of adapting the content (i.e., its format, resolution, bit-rate, frame rate and coefficient dropping) to suit the usage environment properties, the usage environment is adapted (i.e., device resolution to accommodate video resolution, bandwidth) to suit to the content properties and hence provide better quality of service to users. Usage environment adaptation may consider the context and the content that can be described using multimedia metadata, such as the MPEG-7 and MPEG-21, respectively [1].

If the objective is to maximize content quality for an individual user by taking into account only user preferences and device constraints but disregarding shared resource constraints, then the adaptation becomes client-centric [2]. For example, client-centric schemes may allocate bandwidth with prioritization by considering constraints placed on content, system and user, may modify device properties such as resolution and processor speed, may allocate memory and CPU to running process, may adapt operating system policies [3], may alter application properties such as the window size of the media player. However, client-centric schemes can only adapt resources they have access to, i.e., client device and application properties. In some cases, this kind of adaptation is not desirable, for example, where there is limited network bandwidth between clients and servers, higher cost of bandwidth usage and unsupported content format.

Server-centric schemes, on the other hard, consider user preferences and device constraints but in relation to shared computational and resource constraints such as available memory and bandwidth. Consequently, server-centric usage environment adaptation aims to maximize computational and network resource usage whilst providing an average quality of service to users, for example, a differential video service [4]. Server-centric schemes are usually deployed on the same servers where content is stored or intermediate proxies.

Client-centric and server-centric schemes manage resources differently and sometimes apply strategies of varying degrees of fairness such in the case where a service allocates bandwidth to users based on the user and the service. Whilst client-centric schemes may consider the client resource, server centric schemes are constrained by the shared resources, such as the server memory and network bandwidth between server and client. Thus, concurrent requests imply management that takes into account fairness handling and optimum usage of shared resources for the best interest of the server and the network rather than maximizing the experience of individual users. Managing these shared resources without considering the latter issues can lead to unwanted statuses in networks, which can affect both the clients and the content server. Consider for example the case of the bandwidth allocation to multiple client requests. Bandwidth misuse can result in a network bottleneck and bad end-user experience.

When computational and network resources are scarce or client devices can not support the target content, neither usage environment adaptation schemes will function efficiently without simultaneous content adaptation [5].

References

1. J.M. Martínez, "MPEG-7 Tools for Universal Multimedia Access," Journal of the American Society for Information Science and Technology, Vol. 58, No. 9, 2007, pp. 1374–1376.
2. A.A. Sofokleous and M.C. Angelides, "Client-Centric Usage Environment Adaptation Using MPEG-21," Journal of Mobile Multimedia, Vol. 2, No. 4, 2006, pp. 297–310.
3 P.J. Teller and S.R. Seelam, "Insights into Providing Dynamic Adaptation of Operating System Policies," ACM SIGOPS Operating Systems Review, Vol. 40, No. 2, 2006, pp. 83–89.
4. M. Bertini, R. Cucchiara, A. Del Bimbo, and A. Prati, "Semantic Adaptation of Sport Videos with User-Centred Performance Analysis," IEEE Transactions on Multimedia, Vol. 8, No. 3, 2006, pp. 433–443.
5. A.A. Sofokleous and M.C. Angelides. "Non-Random Selection of a Video Content Adaptation Strategy from a Pareto Front," The Computer Journal, forthcoming.

Usage Histories

► Content-user Gap in MPEG-7

User Input

► Interactivity in Multimedia Documents and Systems

User Interaction

► Interactivity in Multimedia Documents and Systems

User Modeling in MPEG-7

Definition

User modeling refers to building a profile of the user's preferences for consumption and usage. In MPEG-7, two tools are specifically implemented for user interaction, which are the User Preferences DS and the User History DS.

Just as we have to model the content to describe the rich multi faceted detail stored both semantically and structurally [1], we must also model the user in a similar fashion, building a profile of the user's preferences for consumption and usage. The requirements are different but the concept remains similar, model the user by capturing the many multi faceted perspectives of the user that combine to describe the mental and corporeal needs of the user. The mental needs are the user's interests and needs for specific information; the corporeal is how, when and where they would like to view that information. The main difference between content modeling and user modeling is that the former is temporally static after creation and the latter evolves continuously over time.

As the user evolves their preferences for information, usage environment and demographic data change as well [2]. The user model must also evolve with the changing user or the model will stagnate. Demographic data does not change a great deal over time and can be adjusted manually. The user's preferences for information and to a lesser extent the usage environment data change at a much greater rate and to update this manually would be impractical. Automatic methods that track and log the user's interaction can do this without any explicit effort from the user. Although the extraction methods can be proprietary or technology specific the logging of the interactions must be MPEG-7 compliant.

Defined in MPEG-7 MDS are two tools that have been specifically implemented for user interaction, namely the User Preferences DS and the User History DS [3]. The User Preferences DS contains all the tools to describe the user both mentally and corporeally by being able to state the preferences for information contained in other tools within MPEG-7 as well the usage environment of the user. The user model consists of three separate components; preferences for content (e.g., movies, books, etc), demographic information (e.g., language, gender, etc) and usage environment (i.e., location, time, etc). The User History DS is the MPEG-7 logging tool for user interactions; it logs the interaction with the user in terms of content viewed, the interaction with the content and where and when interactions took place. The User History DS is then analyzed and used to update the user model, keeping it current to a user's changing needs. User Preferences DS and the User History DS tools can be customized to include or exclude features of the user to provide a granular detail of the user's preferences and consumption requirements. This provides the same amount of rich and perceptively textured description of the user as you would have for content but with the additional functionality of evolving the model as the users needs change.

Cross-References

► Multimedia Content Modeling and Personalization

References

1. I.E.G. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia," John Wiley, New York, 2005.
2. B.L. Tseng, C.-Y. Lin, and J.R. Smith, "Using MPEG-7 and MPEG-21 for Personalizing Video," IEEE Multimedia, Vol. 11, No. 4, October 2004, pp. 42–53.
3. H. Kosch, "Distributed multimedia database technologies supported by MPEG-7 and MPEG-21," CRC Press, Boca Raton, Florida, 2003.

User Preferences

► Content-user Gap in MPEG-7

Utility Model-Based Adaptation of Multimedia Content

Martin Prangl[1], Hermann Hellwagner[1], Tibor Szkaliczki[2]
[1]Klagenfurt University, Klagenfurt, Austria
[2]Hungarian Academy of Sciences, Budapest, Hungary

Synonyms

► Cross-modal utility models; ► Adaptation decision taking engine

Definition

Utility model-based adaptation is used to provide the best possible experience to the user consuming multimedia content under given user preferences, device, and network constraints.

Introduction

Today, multimedia content must be distributed to different devices such as desktop computers, PDAs, and mobile phones. In many delivery situations, the clients are unable to receive and process the content in original quality because of resource limitations, e.g., limited network throughput. One approach to deal with this problem is to employ adaptation of the multimedia content to the actual usage context in order to comply with the given device/network capabilities and constraints. Under the vision of Universal Multimedia Access (UMA) [1], such adaptive multimedia systems have for several years attracted significant research efforts. Typical adaptive multimedia frameworks to date try to consider the technical capabilities of the user's device and possibly the delivery networks' characteristics [2,3], but fail to take into account the user's preferences or the utility of the content for the user. However, the question "How to adapt multimedia data in order to provide the best user perceived utility?" is of central relevance as well, which eventually leads to pursuing a vision of Universal Multimedia Experience (UME) [4].

In addressing this question, i.e., in order to optimize the quality of the adaptation and of the media delivered to the user, the type and information content of the media have to be considered in addition to the technical properties indicated above. For example, in case of an action video, it would most probably be preferable to adapt the video in the spatial domain rather than in the temporal domain [5]; this would result in a smaller video window, but the user would still be able to fully enjoy rapid motion in action scenes. In contrast, in case of a documentary, temporal adaptation (frame rate reduction) may be preferable over spatial adaptation. In addition to these quality aspects, a utility-based adaptation system should optimize the semantic value of the content for the user under the given resource limitations, i.e., satisfy the user's information or entertainment needs.

In this article, a multimedia adaptation decision model is described which uses detailed perceptual quality information and semantic quality estimation. When considering quality in the multimedia area, a perceptual part and a semantic part have to be distinguished [6]. The *perceptual quality (PQ)* is a metric about how a user perceives the content, and refers to the human visual system; for example, a smooth video has a higher perceptual quality than a flickering one. The *semantic quality (SQ)* on the other hand refers to how well the designated information of the media is conveyed to the user, e.g., the semantic content of a news report, or to how well the media consumption entertains the user, e.g., by presenting the full motion of an action video [7]. Furthermore, in this article the term *utility (U)* of a given multimedia content is defined as a metric of the overall satisfaction of the end user consuming this content, resulting from a combination of the perceptual and semantic parts of quality for the given content.

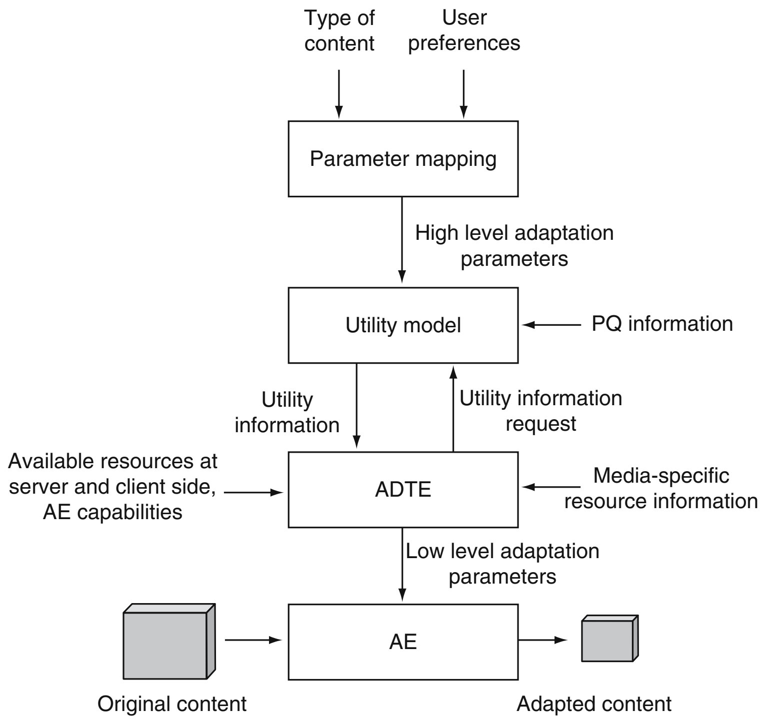
So called *cross-modal utility models* [8] are used to estimate the total utility of a media stream consisting of two or more modalities, e.g., video and audio. The total utility can be expressed as a function which combines the uni-modal utilities of the elementary streams. In the literature, there are some implementations of such a function, discussed in [9], for instance. These implementations rely on adding the weighted uni-modal perceptual qualities, a multiplicative term (multiplication of uni-modal qualities), and specific constants in order to – a posteriori – fit the subjective impressions of a group of test persons. A detailed analysis of such an approach [9] suggests that the implementation of the model itself as well as the weights and constants are strongly dependent on the genre and the subjects participating in the test. In other words, a more generic model for estimating the total audio-visual utility of multimedia content, which can be used for any genre and which takes into account the consumer's preferences, seems to be missing.

Such a utility model, a utility-based multimedia adaptation framework, and an automatic model configuration approach using a recommender strategy are outlined in this article. The adaptation decision taking process involved in this framework, i.e., finding the combination of (adapted) elementary streams of a given content, which complies with the resource constraints and which provides the best audio-visual utility for the consumer, is described as an optimization problem. Four different algorithms for solving this challenging optimization problem, i.e., for finding the "best" adaptation of a given content for a given user within a reasonable (non-annoying) time frame, are presented. Details of the approach outlined in this contribution are given in [10].

Utility-Based Multimedia Framework

In Fig. 1, the concept of the proposed framework with audio-visual (A/V) utility modeling is shown. The preferences of the consumer requesting content and the genre (type) of the requested content (influencing *SQ*) have to be known for configuring the generic *utility model* [11] which is used by the *adaptation decision taking engine (ADTE)* [12]. Currently, five main genre categories are distinguished in a prototype system: action, news, cartoon, documentary, and sports. This input information is mapped (by the *parameter mapping* module) to specific utility model parameters which are called *high level adaptation parameters*, discussed in the next section.

The utility model configured so far additionally needs to be provided the *PQ* of all deliverable content variations. The video variations are characterized by the values of their spatial resolutions, frame rates, and SNR variations. The audio variations are described by the values of their bit rates, sample rates, and number of audio channels. Based on the genre and *PQ* information, the total utility *U* of all deliverable A/V



Utility Model-Based Adaptation of Multimedia Content. Figure 1. Overview of utility-based multimedia framework.

variations can be estimated. Using the utility and the information about the media-specific required resources (e.g., bit rate required for transmission) of each deliverable A/V variation, as well as the information about the available resources on the client and server sides (e.g., the available bandwidth, battery status, or processing power), the ADTE is then able to determine the optimum adaptation strategy for the given content request [12]. This decision has to be taken quickly in order to avoid annoying media startup delays.

The optimum adaptation decision determined is expressed by a set of parameters which are called *low level adaptation parameters*. They define an A/V media stream variation by its features (e.g., frame rate, spatial resolution, sample rate). Based on these target features, the *adaptation engine (AE)* performs the adaptation step on the original content. Finally, the produced variation, fitting the user's preferences and the usage environment and providing the best possible utility under the given conditions, can be delivered for consumption to the requesting client. Note that it is not possible that the ADTE selects a variation that the AE cannot produce because the ADTE has information about the AE capabilities as well.

Utility Model

As already mentioned, the aim of the utility model is to express the user satisfaction of (degraded) media contents. The cross-modal utility model presented in [11] is being used. It expresses the total utility U as a linear combination of the utilities of the audio and the video streams, where in a similar way the utility of an elementary stream can be described as the weighted sum of PQ and SQ. The well-known PSNR metric is used for video PQ estimation. For audio PQ estimation, the common PEAQ metric is applied. The SQ part is modeled as a weighted sum of the relative values of the elementary stream features, where the weights define the relative importance of each stream feature for the individual use case. The SQ definition of the audio stream is given as follows:

$$SQ_A = w_{Sa}\frac{sr}{sr_{orig}} + w_{Ba}\frac{abr}{abr_{orig}} + w_{Ca}\frac{achan}{achan_{orig}}$$
$$w_{Sa}, w_{Ba}, w_{Ca} \in [0,1]$$
$$w_{Sa} + w_{Ba} + w_{Ca} = 1$$
$$sr \leq sr_{orig}, abr \leq abr_{orig}, achan \leq achan_{orig}$$

where sr and sr_{orig} represent the sample rate of the degraded audio variation and the original audio stream, respectively. In a similar manner, *abr* and *achan* represent the encoding audio bit rate and the number of provided audio channels, respectively. All unique stream features defining the degraded stream variation form the low level adaptation parameters (Fig. 1). The importance weights of the stream features, which are called high level parameters, are denoted as w_{Sa}, w_{Ba}, and w_{Ca}, respectively.

The expression for *SQ* of the video stream is given as follows:

$$\begin{aligned} SQ_V = {} & w_{Fv}\frac{fr}{fr_{orig}} + w_{Sv}\frac{height}{height_{orig}}\frac{width}{width_{orig}} \\ & + w_{Qv}\left(1-\frac{q-q_{\min}}{q_{\max}-q_{\min}}\right) \; w_{Fv}, w_{Sv}, w_{Qv} \in [0,1] \\ & w_{Fv}+w_{Sv}+w_{Qv}=1 \\ & fr \le fr_{orig}, height \le height_{orig}, width \le width_{orig} \\ & q \in [q_{\min}, q_{\max}] \end{aligned}$$

where *fr, height, width,* and *q* represent the frame rate, the spatial resolution, and the quantization parameter of the (degraded) video variation, respectively. q_{min} and q_{max} represent the codec (or AE) specific minimum and maximum quantization values. fr_{orig}, $height_{orig}$, and $width_{orig}$ are constants expressing the corresponding feature values of the original video stream. w_{Fv}, w_{Sv}, and w_{Qv} are again importance weight parameters for the stream features.

All high level adaptation parameters have to be adjusted with respect to the given use case. For example, in case of an action video where motion has high priority, a degraded video variation with a high frame rate results in a higher utility than other variations which require equal resources but provide, e.g., higher spatial resolutions and lower frame rates. Well defined high level parameters increase the semantic experience of the consumer.

However, setting the model parameters based on intuitive, hand-crafted rules may not be valid in general. The question is: Are such rules valid for a specific user? Subjective MOS estimation results [10] as well as related subjective perceptual quality tests [13] show that users have different tastes. Thus, applying the same parameters for each user would lead to different multimedia experiences for the individuals. However, the aim of the utility-based multimedia framework is to offer a personalized version of the content that leads to the optimum utility for the individual requesting the content. As already mentioned, the framework takes the individual user preferences into consideration as well. Asking the users for their preferences, i.e., to indicate the high level parameters for the model, would be the obvious way to go. However, in general the users are not experts in the multimedia domain and do not know the optimum settings for the requested content and the actual usage environment. Furthermore, the user would get annoyed if he/she had to answer too many questions in order to provide this helpful information for the system.

For this reason, it would be beneficial for the user, if he/she got a recommendation from users who consumed the same type of content under similar usage environment conditions. In order to achieve such functionality, a *recommender system* was developed for the framework. Its task is to configure the utility model according to the individual user characteristics as well as his/her usage environment. The underlying assumption of this approach is that people who agreed in the past tend to agree again in the future. Information about the content (title, genre), the type of device and its capabilities, available resources as well as demographic information about the user, is taken as input for the recommender system. Together with user satisfaction feedback (which is collected from the user after content consumption) of previous and similar requests, the optimum model parameter setting is predicted. Experiments showed that the success of the recommendation with respect to user satisfaction is increasing with the number of user feedback ratings, which indicates that the recommender system successfully learns. For more information about this automatic parameter setting approach, the reader is referred to [14], where a detailed evaluation is given as well.

Optimization Problem Model of Adaptation Decision Taking

Based on the proposed utility model, which is individually configured for the user, the ADTE has to choose the most appropriate adaptation parameters with respect to the actual resource limitations, device capabilities, user preferences, and AE capabilities, e.g., codec types or quantization levels, in order to provide the maximum media utility to the user. This process has to be fast in order to keep the startup delay of the requested session in a non-annoying range. In particular under dynamic resource limitations, e.g., network bandwidth fluctuations, this decision has to be taken in real-time in order to provide continuous media delivery to the client. In this section, the optimization problem model derived from the above utility model is presented.

A client is requesting a movie m from the streaming media server. The original movie consists of a video and an audio stream. Both the video and audio streams can be adapted into uniquely defined variations, characterized by a set of video features F_v and a set of audio features F_a. They together form the feature set of a movie, denoted by F_m, which can describe the variations of the movie: $F_m = F_v \cup F_a$. The features can be, e.g., spatial resolution, frame rate, type of codec, number of audio channels, and audio sampling rate. Let f_1, $f_2, \ldots, f_n$ denote the features ($n = |F_m|$).

Let V_v and V_a denote the sets of deliverable video and audio variations of movie m on the server (w.r.t. the AE capabilities), respectively. Let V_m denote the set of deliverable variations of movie m, and M and N denote the number of the different video and audio variations, respectively: $M = |V_v|$, $N = |V_a|$. The video and audio streams can be combined arbitrarily into a movie, that is, $V_m = V_v \cdot V_a$. $m|_f$ denotes the value of the feature f of stream m. The particular movie, video and audio variations are denoted by v_m, v_v and v_a, respectively. The variations can be specified as vectors in the feature space. Let A_v and A_a denote the set of video and audio variations, respectively, that the client accepts. It is assumed that both the sets of the deliverable and acceptable variants are discrete and finite. $l(f_k)$ is the number of different available and acceptable values of feature f_k.

Furthermore, the utilities of each deliverable video and audio variation are known or can be calculated. Let $U_V(v_v)$ and $U_A(v_a)$ denote the utilities for video variation $v_v \in V_v$ and audio variation $v_a \in V_a$, respectively. The utility of the multimedia stream resulting from the combination of the video and the audio streams can be calculated as a weighted sum of the utilities of the two modalities: $U(v_m) = (1-\alpha) \cdot U_V(v_v) + \alpha \cdot U_A(v_a)$ where $\alpha \in [0,1]$ denotes the importance weight of the audio utility; α is a high level parameter of the utility model.

Let r denote the number of different resources (processor clock cycles required for encoding and decoding, bit rate). Its value is typically 2 or 3. It is assumed that the resource needs denoted by $C_i(v)$ of each content variation v are known as well ($i = 1, \ldots, r$). Trivially, $C_i(v_m) = C_i(v_v) + C_i(v_a)$. Furthermore, the resources such as processor usage and the total bit rate of the processed streams are limited on the server or the client. Let L_i denote the maximum values of these resources.

The aim is to select a video and an audio variation that the AE is able to produce and such that each of the target features of the multimedia stream satisfies the client request (Eq. 2). The resource needs have to be considered (Eq. 3) and the utility of the multimedia stream resulting from their combination has to be maximized (Eq. 1).

Input:

Client request: $A_v \subset F_v$, $A_a \subset F_a$

Variations on the server (AE specific): $V_v \subset F_v$, $V_a \subset F_a$,

Limits on bandwidth and processor usage: L_i, $i = 1, \ldots, r$

Output:

Movie variation $v_m = (v_v, v_a)$

Maximize

$$\mathrm{U(v_m) = (1-\alpha)\cdot U_V(v_v) + \alpha \cdot U_A(v_a)} \tag{1}$$

Subject to

$$v_v \in V_v \cap A_v, v_a \in V_a \cap A_a \tag{2}$$

$$\mathrm{C_i(v_m) = C_i(v_v) + C_i(v_a) \le L_i,\ i = 1, \ldots, r.} \tag{3}$$

It can be assumed for most of the features that the resource needs as well as the utility are monotonically increasing while the value of a feature is increasing and the other feature values remain unchanged: $(v_1|_f \ge v_2|_f) \Rightarrow U(v_1) \ge U(v_2), C_i(v_1) \ge C_i(v_2)$. This is usually true for each video and audio parameter except the video and audio codec type.

Algorithms to Solve the Optimization Problem

As mentioned earlier, short execution time is crucial in solving the adaptation decision taking (i.e., optimization) problem. Mukherjee et al. [15] gives an overview of the adaptation decision taking process. That paper recommends the total enumeration when the possible feature values are discrete, which is the case in the problem model introduced above. However, the run time can be decreased by exploiting the special characteristics of the problem. In this section, a survey on the algorithms that were implemented and tested is given. For more details on these algorithms, the reader is referred to [10] and [12].

All Combinations

This simple approach checks all combinations of the audio and video variations to find the optimum one. Generating all combinations (i.e., total enumeration) was implemented in order to validate the results of the other algorithms. The time complexity of the algorithm is $T = O(M \cdot N)$.

Merging Video and Audio Variations

This algorithm proceeds with video variations according to the increasing order of bandwidth demand while the audio variations are processed in decreasing order. The time reduction of the algorithm is based on the idea that it is enough to combine the current video variation only with the audio variation of the highest utility among those whose resource needs are less than the available resources minus the video resource need. For efficiency, the ordered list of candidate audio variations is stored in a so called red-black tree, which is a special balanced tree. The algorithm can be efficiently implemented if the number of different resources is at most two.

In general, this method can be used for finding the maximum of a nonlinear global optimization problem which is separable into two groups, that is, the profit (utility) function and the constraints can be written as weighted sums of functions depending on two separate variable groups. The time complexity of the algorithm is $T = O\ (M \cdot \log M + N \cdot \log N)$. This can be reduced to $O\ ((M + N) \cdot \log N)$ if the video variations are ordered in advance according to their bandwidth needs.

Border Scan

This algorithm exploits the monotonicity of the utility and resource needs in the feature values. There are several methods for optimization where the goal function as well as the constraints are monotonic [16]. From the monotonicity of the resource needs and utilities it follows that a point representing the optimum movie variation is located directly below the surface (or border) of the resource constraints in the joint feature space of all modalities. A method is applied where all the points are enumerated and compared with each other that are located directly below the border. The algorithm takes one initial point, then it considers the different features one after the other and extends the border in the subspace of the features examined so far. The time complexity of this border scan algorithm is $T = O\ (M \cdot N/\min(l(f_1), l(f_n)))$ where f_1 and f_n denote the features that the algorithm examined first and last, respectively.

Hill Climbing

Due to the monotonicity in the resource needs and utility, a heuristic search method can be used as well, namely steepest-ascent hill climbing [17]; this was found as an efficient approach for the real-time application at hand. The worst variation is used as a starting point. In each iteration step, the value of that monotonic feature is increased where the utility increase is the highest and the improved variation still satisfies the resource constraints. This algorithm does not necessarily find the optimum because it may run into a local minimum at the border defined by the constraints but it is clearly the fastest algorithm in practical cases. The time complexity of the algorithm is $T = O(\max_i(l(f_i)))$. The accuracy of the algorithm can be significantly improved by starting the algorithm from different initial points and then selecting the best variation from the results of different runs.

Performance Results

The above algorithms were implemented and evaluated on real multimedia stream data [10,12]. Different optimization tasks were generated by varying the high level parameters and resource constraints. Furthermore, a method was examined in which only the spatial resolution of the video was reduced until the result did comply with the resource constraints, at which point the utility was calculated. This latter method is included in order to show how much improvement can be achieved by the utility based adaptation relative to a traditional method which neglects the utility aspects. Each recommended algorithm was much better than reducing the spatial resolution of the video stream only, leaving other stream parameters untouched. Clearly, heuristic search (hill climbing) was the fastest but it does not always find the exact optimum. It was observed that this algorithm is more likely to fail in finding the optimum if the resource limits are low; its quality could be improved by repeating it from different initial points. Generating all combinations was not too inefficient because the number of audio variations was small in these experiments. Merging was the slowest but its run time can be highly reduced if sorting is done in advance before client requests arrive.

Conclusion

In this article, a generic audio-visual utility model for multimedia content adaptation was presented that is able to consider the user preferences and the usage environment as well as different content genres. For efficient parameter setting, a recommender-based approach was introduced that configures the model to the user's individual utility notion based on intuitive rules, users' judgments, demographic features, and favorite content types. Finding video and audio stream variations that maximize the media stream's utility (or the experience) for the user based on the proposed model under given resource constraints, represents a complex optimization problem in the multimedia area. Four algorithms to find optimum video and audio variations for multimedia content adaptation were presented, implemented, and evaluated. The simple heuristic hill-climbing optimization method was found to be the most efficient. However, this algorithm may fail to find the optimum, so it has to be used with care and potentially has to be improved. The merging method is recommended especially when the utility function is non-monotonic and preparation (sorting) can be done before client requests arrive. Border scan is efficient in the monotonic case if the hill climbing approach fails.

Applying the presented approach to an adaptive multimedia framework yields a better multimedia experience for the client.

References

1. A. Vetro, C. Christopoulos, and T. Ebrahimi, (Eds.), "Special Issue on Universal Multimedia Access," IEEE Signal Processing Magazine, Vol. 20, No. 2, March 2003.
2. D. Jannach, K. Leopold, C. Timmerer, and H. Hellwagner, "A Knowledge-Based Framework for Multimedia Adaptation," Applied Intelligence, Vol. 24, No. 2, April 2006, pp. 109–125.
3. F. López, J.M. Martínez, and V. Valdés, "Multimedia Content Adaptation within the CAIN Framework via Constraints Satisfaction and Optimization," Proceedings of the Fourth International Workshop on Adaptive Multimedia Retrieval (AMR) – Springer LNCS 3877, Geneva, Switzerland, July 2006.
4. F. Pereira and I. Burnett, "Universal Multimedia Experiences for Tomorrow," IEEE Signal Processing Magazine, Vol. 20, No. 2, Mar. 2003, pp. 63–73.
5. H. Knoche, J.D. McCarthy, and M.A. Sasse, "Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV," Proceedings of ACM Multimedia, pp. 829–838, November 2005.
6. T.C. Thang, Y.J. Jung, and Y.M. Ro, "Modality Conversion for QoS Management in Universal Multimedia Access," IEE Proceedings – Vision, Image, and Signal Processing, Vol. 152, June 2005, pp. 374–384.
7. T.C. Thang, Y.J. Jung, and Y.M. Ro, "Semantic Quality for Content-Aware Video Adaptation," Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP'05), pp. 1–4, October 2005.
8. S. Winkler and C. Faller, "Audiovisual Quality Evaluation of Low-Bitrate Video," Proceedings of the SPIE/IS&T Human Vision and Electronic Imaging, Vol. 5666, January 2005, pp. 139–148.
9. D.S. Hands, "A Basic Multimedia Quality Model," IEEE Transactions on Multimedia, Vol. 6, No. 6, Dec. 2004, pp. 806–816.
10. M. Prangl, T. Szkaliczki, and H. Hellwagner, "A Framework for Utility-based Multimedia Adaptation," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 6, June 2007, pp. 719–728.
11. M. Prangl, H. Hellwagner, and T. Szkaliczki, "A Semantic-Based Multi-Modal Utility Approach for Multimedia Adaptation," Proceedings of the Seventh International Workshop on Image Analysis for Multimedia Services (WIAMIS), pp. 67–70, April 2006.
12. M. Prangl, H. Hellwagner, and T. Szkaliczki, "Fast Adaptation Decision Taking for Cross-modal Multimedia Content Adaptation," Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 137–140, July 2006.
13. ITU-R BT500–11, Methodology for the Subjective Assessment of the Quality of Television pictures, 2002.
14. M. Prangl, R. Bachlechner, and H. Hellwagner, "A Hybrid Recommender Strategy for Personalised Utility-Based Cross-Modal Multimedia Adaptation," Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1707–1710, July 2007.
15. D. Mukherjee, E. Delfosse, J-G. Kim, and Y. Wang, "Optimal Adaptation Decision-Taking for Terminal and Network Quality of Service," IEEE Transactions on Multimedia, Vol. 7, No. 3, June 2005, pp. 454–462.
16. H. Tuy, "Monotonic Optimization: Problems and Solution Approaches," SIAM Journal on Optimization, Vol. 11, No. 2, 2000, pp. 464–494.
17. E. Rich and K. Knight, Artificial Intelligence, MacGraw-Hill, 1991.